DOCUMENT RESUME

ED 194 534 TM 800 660

AUTHOR Tatsuoka, Kikumi
TITLE The Least-Squares Estimation of Latent Trait Variables.
PUB DATE Apr 80
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Computer Assisted Testing; *Error of Measurement; Higher Education; *Latent Trait Theory; *Least Squares Statistics; Mathematical Models; *Maximum Likelihood Statistics; *Multiple Regression Analysis; *Scoring Formulas; Test Theory

ABSTRACT

This paper presents a new method for estimating a given latent trait variable by the least-squares approach. The beta weights are obtained recursively with the help of Fourier series and expressed as functions of item parameters of response curves. The values of the latent trait variable estimated by this method and by maximum likelihood method were compared using real data: A 48-item test in matrix algebra administered by computer, in an adaptive testing situation. The results were very close, and yet the values of the latent trait variable estimated by the multiple regression method were always obtainable, even for a small number of examinees and a small number of items. The maximum likelihood method, on the other hand, often fails to converge. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194534

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE LEAST-SQUARES ESTIMATION OF LATENT TRAIT VARIABLES

by

Kikumi Tatsuoka

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. Tatsuoka

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Computer-based Education Research Laboratory
University of Illinois at Urbana-Champaign

Presented at the 1980 Annual Meeting of AERA

TM 800 660

THE LEAST-SQUARES ESTIMATION OF LATENT TRAIT VARIABLES

ABSTRACT

This paper presents a new method for estimating a given latent trait variable by the least-squares approach. The beta weights are obtained recursively with the help of Fourier series and expressed as functions of item parameters of response curves. The values of the latent trait variable estimated by this method and by maximum likelihood method were compared using real data. The results were very close, and yet the values of the latent trait variable estimated by the multiple regression method were always obtainable. The maximum likelihood method, on the other hand, often fails to converge.

INTRODUCTION

Item Characteristic Curve Theory (ICCT)[1] is a test theory model that has been utilized for assessing an individual student's ability or achievement level, θ, in a tailored testing[2] situation where each student takes a different set of items. The item parameters of a large pool are previously estimated and they are used to estimate ability level (Lord, 1970; Weiss, 1976).

Methods of estimating parameters have been investigated by Bock (1972), Jensema (1976), Lord (1975), Samejima (1978), and Urry (1974). Two estimative methods, the maximum likelihood (MXL) procedure and the Urry method, have been commonly used and their computer programs are known as LOGIST (Wood, Wingersky & Lord, 1976) and ANCILLES (developed at the U.S. Civil Service Comission). Swaminathan and Gifford (1979) compared the two methods as to the accuracy of estimation and the effect of violating underlying assumptions. The results of their Monte Carlo study show that the MXL procedure is superior to the Urry procedure when test items and ability levels are estimated simultaneously. But both methods yielded poor estimates of discrimination parameters for fifteen items or less. Moreover, it is customarily known that both the methods, especially the MXL method, require a substantial number of subjects in the estimation procedure.

These drawbacks cause a serious problem when ICCT models are considered for use in conjunction with criterion-referenced testing,

or achievement tests that are used as an integral part of instruction. Tests administered in these cases are usually short, and moreover, it is not always easy to collect a large amount of valid data for parameter estimation.

However, estimating the ability parameter, θ, from a set of known item parameters, instead of estimating both ability and item parameters simultaneously, is easily done by using the MXL procedure (Lord and Novick, 1978) or Owen's updating Bayesian method (1975). The MXL method provides, in general, better estimates than the Owen method, but it is impossible to give a finite estimate for a set of responses in which all items are answered right or wrong. These special cases often happen at the beginning of the adaptive testing procedure. Fortunately, the Owen method provides estimates for such response patterns. Therefore, the Owen method is often used for assessing ability level at the initial stage of adaptive testing and then replaced by the MXL method later (Brown and Weiss, 1977; Kingsbury and Weiss, 1979). The Owen method tends to depend on the sequential order of items to be presented (Sympson, 1977; Bejar and Weiss, 1979).

This study proposes an alternative estimation procedure of ability or achievement level, θ, with or without using a set of known item parameters. The estimate of $\underset{\sim}{\theta}$ is obtained by minimizing the distance between a variable $\underset{\sim}{\theta}$ and the projection of $\underset{\sim}{\theta}$ onto the space spanned by observed item-scores. In other words, the estimate of $\underset{\sim}{\theta}$ is expressed as a multiple regression equation of $\underset{\sim}{\theta}$ as a criterion variable and

observed item scores as predictors where each weight β is a function of the correlations between $\underset{\sim}{\theta}$ and each item i. The estimate of $\underset{\sim}{\theta}$ is given by a function of intraclass correlation coefficients $\rho_{\theta i}$ which, in turn, can be expressed as functions of item discriminating powers (a) and difficulties (b) obtained from ICCT models (Lord and Novick, 1968). If each item discriminating power is set to unity, then it should provide estimates of $\underset{\sim}{\theta}$ for the one-parameter logistic or Rasch model. Further investigation of comparisons among the above-mentioned cases and characteristics of this least-squares model is left for future research. In this paper, the least-squares and MXL methods are compared with respect to the estimates of $\underset{\sim}{\theta}$ with known item parameters of the two-parameter logistic model. The data used in the comparison are obtained from a 48-item matrix algebra test administered to students in a multivariate statistics course at the University of Illinois.

MULTIPLE REGRESSION FUNCTIONS AND PROJECTIONS

Person space: Suppose $\underset{\sim}{X}_g$ is the raw-score vector of a given item g in a population, or more precisely a space spanned by the persons in the population. In Item Characteristic Curve Theory (or latent trait theory), a response pattern for person I is crucial for determining the likelihood function, but the focus of our interest in this paper will be switched to the score vector of item g in the person-space where the two-parameter ICCT model is considered. Let us consider a deviation score vector $\underset{\sim}{x}_g = (X_{g1}, X_{g2}, \ldots X_{gN}) - (\mu_g, \mu_g, \ldots, \mu_g)$ for convenience in the sequel.

If the norm of $\underset{\sim}{x}_g$ is defined by

$$||\underset{\sim}{x}_g|| = \left[\sum_{i=1}^{N} x_{gi}x_{gi}\right]^{1/2},$$

then the standard deviation of item g becomes the norm of $\underset{\sim}{x}_g$ divided by $(N-1)^{1/2}$, and the covariance of two vectors $\underset{\sim}{x}_g$ and $\underset{\sim}{x}_f$ is the inner product of these two vectors divided by N-1. Two vectors $\underset{\sim}{x}_g$ and $\underset{\sim}{x}_f$ are said to be orthogonal if their inner product equals zero, that is $(\underset{\sim}{x}_g, \underset{\sim}{x}_f) = 0$. A vector $\underset{\sim}{x}_g$ is said to be orthogonal to a set M if $\underset{\sim}{x}_g$ is orthogonal to every vector in M. A vector space with these properties (direct consequences of the definition of the inner product of any two vectors), provides a wealth of structural characteristics that imply very useful analytical results applicable to many problems.

For example, the shortest vector from a point to a subspace M in n-dimensional Euclidean space is orthogonal to the subspace M. This is a special case of the optimization principle which is called the projection theorem. Tatsuoka (1975) discussed in detail the relationship among projection operators, regression functions and multiple regression functions, so only a brief exposition will be given here.

The projection theorem can be explained briefly as follows: For a given vector $\underset{\sim}{x}$ and a subspace M, there exists a unique vector $\underset{\sim}{m}$ in M which is closest to $\underset{\sim}{x}$ in the sense that it minimizes $||\underset{\sim}{x} - \underset{\sim}{m}||$. Furthermore a necessary and sufficient condition that $\underset{\sim}{m}$ be the unique minimizing vector is that $\underset{\sim}{x} - \underset{\sim}{m}$ be orthogonal to M. Hence, $\underset{\sim}{m}$ can be called the multiple regression function $R(\underset{\sim}{x} \mid M)$ and the residual will be $\underset{\sim}{x} - \underset{\sim}{m}$ which is orthogonal to $\underset{\sim}{m} = R(\underset{\sim}{x} \mid M)$. If M consists of a single vector, $\underset{\sim}{y}$, then $R(\underset{\sim}{x} \mid M)$ is the regression function of $\underset{\sim}{x}$ onto $\underset{\sim}{y}$. It is always true that $R(\underset{\sim}{x} \mid M)$ is linear and idempotent; that is for any real numbers s and t,

$$R(s\underset{\sim}{x}_g + t\underset{\sim}{x}_f \mid M) = sR(\underset{\sim}{x}_g \mid M) + tR(\underset{\sim}{x}_f \mid M)$$

$$R(R(\underset{\sim}{x}_g \mid M) \mid M) = R(\underset{\sim}{x}_g \mid M).$$

Normal equations in the context of Euclidean Geometry: Our concern in this study is to express the multiple regression of $\underset{\sim}{\theta}$ onto $\{\underset{\sim}{x}_i\}$ as a linear combination of the $\underset{\sim}{x}_i$, $\hat{\underset{\sim}{\theta}} = d_1\underset{\sim}{x}_1 + d_2\underset{\sim}{x}_2 + \ldots + d_n\underset{\sim}{x}_n$, equivalent to that of finding the n scalars d_i, $i = 1, \ldots, n$, that minimize expression (1).

$$(1) \quad ||\underset{\sim}{\theta} - d_1\underset{\sim}{x}_1 - d_2\underset{\sim}{x}_2 - \ldots - d_n\underset{\sim}{x}_n||$$

or, setting the inner product equal to zero,

$$(2) \quad (\underset{\sim}{\theta} - d_1\underset{\sim}{x}_1 - d_2\underset{\sim}{x}_2 - \ldots - d_n\underset{\sim}{x}_n, \underset{\sim}{x}_i) = 0$$

for each i. By rewriting Equation (2), we obtain Equation (3)

$$\begin{aligned}
&d_1(\underset{\sim}{x}_1,\underset{\sim}{x}_1) + d_2(\underset{\sim}{x}_2,\underset{\sim}{x}_1) + \ldots + d_n(\underset{\sim}{x}_n,\underset{\sim}{x}_1) = (\underset{\sim}{\theta},\underset{\sim}{x}_1)\\
&d_1(\underset{\sim}{x}_1,\underset{\sim}{x}_2) + d_2(\underset{\sim}{x}_2,\underset{\sim}{x}_2) + \ldots + d_n(\underset{\sim}{x}_n,\underset{\sim}{x}_2) = (\underset{\sim}{\theta},\underset{\sim}{x}_2) \qquad (3)\\
&\qquad\qquad \ldots\ldots\ldots\ldots\ldots\ldots\\
&d_1(\underset{\sim}{x}_1,\underset{\sim}{x}_n) + d_2(\underset{\sim}{x}_2,\underset{\sim}{x}_n) + \ldots + d_n(\underset{\sim}{x}_n,\underset{\sim}{x}_n) = (\underset{\sim}{\theta},\underset{\sim}{x}_n)
\end{aligned}$$

This set of n equations with n coefficients d_i is known as the normal equations for the least-squares problems if $\underset{\sim}{\theta}$ is replaced by any criterion vector. If the n x n matrix of $(\underset{\sim}{x}_i,\underset{\sim}{x}_j)$, $i,j = 1, \ldots, n$ has a non-zero determinant, then Equation (3) can be solved for the coefficients d_i. However, since $\underset{\sim}{\theta}$ is a nonobservable variable, the straightforward application of solving n simultaneous equations will not work in this case. Moreover, the evaluation of an n x n determinant

is not an easy task, especially if n is large. As an alternative approach, we will introduce an analytical method using Fourier series to estimate $\underset{\sim}{\theta}$.

Gram-Schmidt procedure: Let us apply the Gram-Schmidt procedure to a set of vectors $\{\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n\}$ (or $\{\underset{\sim}{x}_i\}$), obtaining an orthonormal set $\{\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_n\}$ as the basis of the space spanned by $\underset{\sim}{x}_i$. Then any vector $\underset{\sim}{y}$ can be approximated by its projection onto this space, which can be written in the form of a Fourier series (Dunford and Schwartz, 1964; Simmons. 1963; Tatsuoka, 1975), as follows:

$$(4) \quad \hat{\underset{\sim}{y}} = \sum_{i=1}^{n} (\underset{\sim}{y}, \underset{\sim}{e}_i)\underset{\sim}{e}_i$$

Since $\underset{\sim}{y} - \hat{\underset{\sim}{y}}$ is orthogonal to the space spanned by $\{\underset{\sim}{e}_i\}$, which is equivalent to saying that $\underset{\sim}{y} - \hat{\underset{\sim}{y}}$ is orthogonal to the space spanned by $\{\underset{\sim}{x}_i\}$, the estimated $\underset{\sim}{y}$ will be easily obtained once the independent vectors $\underset{\sim}{x}_i$ are orthogonalized. Equation (4) relates closely with the Gram-Schmidt procedure because the numerator of Equation (5) below is the residual vector $\underset{\sim}{x}_i - \hat{\underset{\sim}{x}}_i$, where $\hat{\underset{\sim}{x}}_i$ is the regression of $\underset{\sim}{x}_i$ onto the space M_{i-1} spanned by $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_{i-1}$ or equivalently by $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_{i-1}$ and moreover $\underset{\sim}{x}_i - \hat{\underset{\sim}{x}}_i$ is orthogonal to space M_{k-1}.

$$\underset{\sim}{e}_i = \frac{\underset{\sim}{x}_i - R(\underset{\sim}{x}_i \mid \underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_{i-1})}{\left\| \underset{\sim}{x}_i - R(\underset{\sim}{x}_i \mid \underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_{i-1}) \right\|}$$

(5)

$$= \frac{\underset{\sim}{x}_i - R(x_i \mid e_1, e_2, \ldots, e_{i-1})}{K_i} \quad ,$$

where $\underset{\sim}{e}_i$, $i = 1, \ldots, n$ are normalized residual vectors while $\underset{\sim}{x}_i - \hat{\underset{\sim}{x}}_i$ are not normalized. If we consider the space spanned by $\{\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n, \theta\}$, adding $\underset{\sim}{\theta}$ to $\{\underset{\sim}{x}_k\}$ then the last orthogonalized vector $\underset{\sim}{e}_{n+1}$ by the Gram-Schmidt procedure has the numerator of $\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}}$, where the estimated $\underset{\sim}{\theta}$ is given by Equation (6) below, replacing vector $\underset{\sim}{y}$ in Equation (4) by $\underset{\sim}{\theta}$

$$\hat{\underset{\sim}{\theta}} = \sum_{i=1}^{n} (\underset{\sim}{\theta}, \underset{\sim}{e}_i)\underset{\sim}{e}_i \tag{6}$$

Also, the Gram-Schmidt process can be said to be a procedure for inverting the matrix given by Equation (3) and can itself be viewed as the solution of a minimum-norm approximation or the least-squares problem.

ESTIMATION OF LATENT VARIABLE θ

Calculation Formula: The estimation of $\underset{\sim}{\theta}$ was given in Equation (6), but note that the Fourier coefficients $(\underset{\sim}{\theta}, \underset{\sim}{e}_i)$, $i = 1, \ldots, n$ include the unknown variable $\underset{\sim}{\theta}$. Therefore they must be approximated by measurable variables. The goal can be achieved by first rewriting $\underset{\sim}{e}_i$ as a linear combination of $\underset{\sim}{x}_i$, obtaining Equation (7).

$$\underset{\sim}{e}_i = \sum_{k=1}^{i} \alpha_{ik} \underset{\sim}{x}_k \tag{7}$$

The coefficients α_{ik} can be determined recursively and given as follows

$$\alpha_{ik} = -\left[\sum_{j=k}^{i-1} (\underset{\sim}{x}_i, \underset{\sim}{e}_j)\alpha_{jk}\right] / K_i \tag{8}$$

Substituting $\underset{\sim}{e}_i$ in Equation (6) by Equation (7), $\hat{\underset{\sim}{\theta}}$ is expressed as a linear combination of $\underset{\sim}{x}_k$.

$$\hat{\underset{\sim}{\theta}} = \sum_{i=1}^{n} \left[\sum_{k=1}^{i} (\underset{\sim}{\theta}, \underset{\sim}{e}_i)\alpha_{ik} \underset{\sim}{x}_k\right] \tag{9}$$

$$= \sum_{k=1}^{n} \left[\sum_{j=k}^{n} (\underset{\sim}{\theta}, \underset{\sim}{e}_j)\alpha_{jk}\right] \underset{\sim}{x}_k$$

where the last member in the above equations is obtained by taking the summation of terms column-wise, while the terms in the second member are summed row-wise.

Substituting $\underset{\sim}{e}_j$ in $(\underset{\sim}{\theta},\underset{\sim}{e}_j)$ by Equation (7), and taking the sumation over i and the coefficients α_{ji} out of the parentheses of the inner product, the middle term of Equation (10) below is obtained. Since $(\underset{\sim}{\theta},\underset{\sim}{x}_k)$ is written as the product of the correlation between $\underset{\sim}{x}_k$ and $\underset{\sim}{\theta}$, $\rho_{\theta k}$, and the standard deviation of $\underset{\sim}{x}_k$, $\sigma(\underset{\sim}{x}_k)$ with the multiplier N-1, thus we obtain the rightmost term of Equation (10) by substituting the relation for $(\underset{\sim}{\theta},\underset{\sim}{x}_k)$.

$$(10) \qquad \sum_{j=k}^{n} (\underset{\sim}{\theta},\underset{\sim}{e}_j)\alpha_{jk} = \sum_{j=k}^{n} \left[\left(\sum_{i=1}^{j} (\underset{\sim}{\theta},\underset{\sim}{x}_i)\alpha_{ji} \right] \alpha_{jk} \right.$$

$$= (N-1) \sum_{j=k}^{n} \left[\sum_{i=1}^{j} \rho_{\theta x_i} \sigma(\underset{\sim}{\theta})\sigma(\underset{\sim}{x}_i)\alpha_{ji} \right] \alpha_{jk}$$

Substituting the rightmost member of Equation 10 in Equation 9 and assuming $\sigma(\underset{\sim}{\theta}) = 1$, the final form for actual calculation is derived as follows:

$$(11) \qquad \hat{\underset{\sim}{\theta}} = (N-1) \sum_{k=1}^{n} \left[\sum_{j=k}^{n} \left(\sum_{i=1}^{j} \rho_{\theta x_i} \sigma(\underset{\sim}{x}_i)\alpha_{ji} \right) \alpha_{jk} \right] x_k .$$

Hence the coefficients of x_k in Equation (11) are expressed by Equation (12).

$$(12) \qquad \beta_k = (N-1) \sum_{j=k}^{n} \sum_{i=1}^{j} \rho_{\theta x_i} \sigma(\underset{\sim}{x}_i)\alpha_{ji}\alpha_{jk} \quad ,$$

or equivalently

$$\beta_k = (N-1) \sum_{i=1}^{n} \rho_{\theta x_i} \sigma_{(\underset{\sim}{x}_i)} \left(\sum_{j=k}^{n} \alpha_{ji} \alpha_{jk} \right).$$

Equation (12) shows that the βs for earlier terms in the multiple regression equation remain unaltered when subsequent terms are added in a stepwise manner.

Note that the x_j are deviation score vectors, hence the right-hand side of Equation (11) is simply a linear combination of $\underset{\sim}{x}_k$ with coefficients β_k, which is an ordinary multiple regression equation with standardized coefficients. Thus

$$(13) \qquad \hat{\underset{\sim}{\theta}} = \sum_{k=1}^{n} \beta_k \underset{\sim}{x}_k = R(\underset{\sim}{\theta} \mid \underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n).$$

Estimation of $\rho_{\theta x_i}$: If we assume that the latent trait θ follows the normal distribution with mean 0 and standard deviation unity, then Lord and Novick (1968) showed an approximation to $\rho_{\theta x_i}$ as a function of item discriminating power a_i and the proportion π_i of correct answers to item i, as follows:

$$(14) \qquad \rho_{\theta x_i} = \frac{a_i}{(1+a_i^2)^{1/2}} \quad \frac{\phi(\gamma_i)}{(\pi_i(1-\pi_i))^{1/2}}$$

where ϕ is the normal density function and γ_i is such a value that $\int_{\gamma_i}^{\infty} \phi(t)dt = \pi_i$. If x_i is continuous, then the approximate $\rho_{\theta x_i}$ is given by Equation (15):

$$\rho_{\theta x_i} = \frac{a_i}{(1+a_i^2)^{1/2}} \tag{15}$$

By substituting the appropriate approximation of $\rho_{\theta x_i}$ in Equations (11) and (12), the β weights and the estimated $\hat{\underset{\sim}{\theta}}$ for each individual will be easily calculated, provided that the a_i are assumed to have been estimated earlier in a large sample by the maximum likelihood estimation method. Note that using previously estimated a's and b's to obtain the estimation of θ's has been commonly practiced in adaptive testing situations.

The Rasch model assumes that all items have the same discriminating powers and thus they can be different only in their difficulties. By setting each a_i to unity (Lord & Novick, 1978) in Equation (14) and (15) Rasch estimates of $\underset{\sim}{\theta}$ can be obtained in the same manner from Equations (11) and (12).

In cases when θ does not follow a normal distribution, one must find another way to approximate $\rho_{\theta x_i}$ for $i = 1, \ldots, n$, either numerically or algebraically.

ACCURACY OF THE ESTIMATION

Standard error of prediction: The standard error of prediction is given by the norm of the residual vector $\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}}$, divided by $(N-1)^{\frac{1}{2}}$. The squared norm of the residual is expressible as the difference between the squared norms of $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$:

$$(16) \qquad ||\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}}||^2 = ||\underset{\sim}{\theta}||^2 - ||\hat{\underset{\sim}{\theta}}||^2$$

(Equation (19), below, is used in showing this to be true.)

Equation (16) describes a relation between the estimated $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$. If the norms of $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$ are equal, then the standard error of prediction will be zero. Also, the latent trait $\underset{\sim}{\theta}$ can never have a smaller norm than the estimated $\hat{\underset{\sim}{\theta}}$, $||\underset{\sim}{\theta}|| \geq ||\hat{\underset{\sim}{\theta}}||$, since the left hand side of Equation (16) is always nonnegative. In other words, $\underset{\sim}{\theta}$ has larger standard deviation than $\hat{\underset{\sim}{\theta}}$.

Multiple R: The correlation of the estimated $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$ itself indicates to what extent $\hat{\underset{\sim}{\theta}}$ represents the unknown vector $\underset{\sim}{\theta}$. In order to show the relationship between these two vectors, some further properties are needed.

The inner product of $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$ is the sum of squared Fourier coefficients:

$$(17) \qquad (\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}) = \sum_{i=1}^{n} (\underset{\sim}{\theta}, \underset{\sim}{e}_i)^2$$

A simple calculation leads us to the relation that the squared norm of $\hat{\underset{\sim}{\theta}}$ is also the sum of squared Fourier coefficients, given as follows:

$$(18) \qquad ||\hat{\underset{\sim}{\theta}}||^2 = \left[\sum_{i=1}^{n} (\underset{\sim}{\theta},\underset{\sim}{e}_i)\underset{\sim}{e}_i, \quad \sum_{i=1}^{n} (\underset{\sim}{\theta},\underset{\sim}{e}_i)\underset{\sim}{e}_i\right] = \sum_{i=1}^{n} (\underset{\sim}{\theta},\underset{\sim}{e}_i)^2$$

since $(\underset{\sim}{e}_i,\underset{\sim}{e}_j) = 0$ for $i \neq j$, $(\underset{\sim}{e}_i,\underset{\sim}{e}_j) = 1$ for $i = j$. Therefore Equation (19) below, stating that the norm of $\hat{\underset{\sim}{\theta}}$ equals the inner product of $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$, is true.

$$(19) \qquad ||\hat{\underset{\sim}{\theta}}||^2 = (\underset{\sim}{\theta},\hat{\underset{\sim}{\theta}})$$

The squared correlation between $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$ is given by

$$(20) \qquad R^2 = \text{corr}^2(\underset{\sim}{\theta},\hat{\underset{\sim}{\theta}}) = \frac{1}{||\underset{\sim}{\theta}||^2} \sum_{i=1}^{n} (\underset{\sim}{\theta},\underset{\sim}{e}_i)^2$$

Meanwhile, the squared correlation between $\underset{\sim}{\theta}$ and $\underset{\sim}{e}_i$ is given by

$$(21) \qquad R_i^2 = \text{corr}^2(\underset{\sim}{\theta},\underset{\sim}{e}_i) = \frac{1}{||\underset{\sim}{\theta}||^2} (\underset{\sim}{\theta},\underset{\sim}{e}_i)^2$$

Therefore, the squared multiple R of the multiple regression $\underset{\sim}{\theta}$ equals the sum of the squared correlation between $\underset{\sim}{\theta}$ and $\underset{\sim}{e}_i$; i.e., Equation (22) holds:

$$R^2 = R_i^2 + \ldots + R_n^2 \qquad (22)$$

Let us change the notation of the component R_i in Equation (22), to $R_{\theta e_i}$ so that the distinction from $R_{\hat{\theta} e_i}$, the correlation between $\hat{\underset{\sim}{\theta}}$ and $\underset{\sim}{e}_i$, will be more explicit. It can be shown that $R_{\hat{\theta} e_i}$ is given by the following formula:

$$R_{\hat{\theta} e_i}^2 = \frac{(\underset{\sim}{\theta}, \underset{\sim}{e}_i)^2}{\Sigma(\underset{\sim}{e}_j, \underset{\sim}{\theta})^2} = \frac{(\underset{\sim}{\theta}, \underset{\sim}{e}_i)^2}{||\hat{\underset{\sim}{\theta}}||^2} \qquad (23)$$

The numerators of $R_{\theta e_j}^2$ and $R_{\hat{\theta} e_j}^2$ are same, but the denominators are the squared norms of θ and $\hat{\theta}$ respectively. It should be noted that the correlation of $\underset{\sim}{\theta}$ and $\underset{\sim}{e}_i$, $R_{\theta e_j}$ is smaller or equal to the correlation of $\hat{\underset{\sim}{\theta}}$ and $\underset{\sim}{e}_j$, $R_{\hat{\theta} e_j}$ because $||\hat{\theta}|| \leq ||\theta||$. That is, $R_{\hat{\theta} e_j} \geq R_{e_j}$, the correlation of estimated $\underset{\sim}{\theta}$ with each $\underset{\sim}{e}_i$ is always larger than or equal to the correlation of $\underset{\sim}{\theta}$ itself and each e_i. $R_{\hat{\theta} e_j}$ will be equal to $R_{\theta e_j}^2$ if and only if $||\hat{\underset{\sim}{\theta}}|| = ||\underset{\sim}{\theta}||$

From Equation (16) and the discussions subsequent to it, it follows that the squared multiple R of $\underset{\sim}{\theta}$ and $\underset{\sim}{e}_i$, $R_{\theta e_i}^2$, is generally smaller than the squared multiple R of $\hat{\underset{\sim}{\theta}}$ with e_i, $R_{\hat{\theta} e_i}^2$. This relation implies that the correlation of $\hat{\underset{\sim}{\theta}}$ with each $\underset{\sim}{e}_i$ tends to provide a bit inflated R value compared with the correlation of the latent trait variable $\underset{\sim}{\theta}$ itself with $\underset{\sim}{e}_i$. The squared correlation of $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$ is expressed as the sum of the squared correlation of $\underset{\sim}{\theta}$ with each element of $\{e_i\}$ according to Equation (22). If we replace each term of Equation (22) by the estimated squared correlation, $R_{\hat{\theta} e_i}^2$,

then the newly obtained estimated $R^2_{\theta\hat{\theta}}$ will be larger than the original squared multiple R. This fact warns us that the estimated R^2 is inflated if we use the estimated $\hat{\underset{\sim}{\theta}}$ instead of $\underset{\sim}{\theta}$ itself.

Finally, the squared multiple R of $\underset{\sim}{\theta}$ and $\hat{\underset{\sim}{\theta}}$ is given by Equation (24) as follows:

$$R^2_{\underset{\sim}{\theta}\hat{\underset{\sim}{\theta}}} = \left[\sum_{i=1}^{n} \sum_{k=1}^{l} \alpha_{ik} \| \underset{\sim}{x}_k \| \rho_{\theta x_i} \right]^2 . \tag{24}$$

Equations (11), (12) and (24) are derived without restrictive assumptions.

If we replace θ by some known variable, say a criterion variable y, then Equations (11), (12) and (24) will be reduced to the case of multiple regression theory.

APPLICATION TO ADAPTIVE TESTING

In some adaptive testing situations, an item is chosen so as to maximize the information function of a set of items (Samejima, 1978). Thus, each examinee takes a different set and number of items according to his/her performance. An examinee's ability level, $\underset{\sim}{\theta}$, can be easily obtained by the MXL method when item parameters are known. However, if an examinee correctly answers several questions given consecutively, which will often happen at the beginning of adaptive testing procedures, then we cannot get a finite estimate of $\underset{\sim}{\theta}$ for him/her. The MXL method is not applicable to responses consisting of either all 1's or 0's (Kendall and Stewart, 1978). Moreover, this method does not work for some unusual response patterns, as can be seen in the following table. The item parameters used in the examples of Table 1 are obtained from the 48-item Stanford Achievement Test vocabulary scale, calibrated by LOGIST using about 2,000 examinees. The order of items in each response pattern (the order of items given in the test) agrees with the order of item difficulties with a few exceptions. Hence, it can be said that the items in the response patterns are pretty much arranged from easiest to hardest. The first two examples are actual response patterns obtained from the vocabulary test, but the rest are hypothetical patterns.

Insert Table 1 about here

Since these examples are for the three-parameter logistic model, they may be more often subject to the problem of nonconvergence of parameter estimation by the MXL method than is the case for the two-parameter logistic model. The proposed model may be expanded to the three-parameter model without much trouble.

The Owen (1975) method always provides estimates of $\underset{\sim}{\theta}$ but the resulting scores are order-dependent (Sympson, 1977). Rearranging the order of items results in slightly different estimates of $\underset{\sim}{\theta}$. Moreover, the posterior distribution after each item is administered is approximated by a normal distribution; thus, inaccuracy of the estimates probably increases as the number of items becomes larger.

By the least-squares method, $\underset{\sim}{\theta}$ is always obtainable and is free from dependence on the item order. But it is subject to the typical problems of multiple regression theory, such as regression toward the mean and the normality assumption of $\underset{\sim}{\theta}$.

The following example shows how the new estimation method works with results of adaptive testing of the matrix algebra test.

Example

A 48-item matrix algebra test was administered to students of a multivariate statistics course at the University of Illinois. The students had taken a series of matrix algebra lessons on PLATO (a computer-based instruction system developed at the University of Illinois). The testing was designed to minimize guessing by permitting students who did not know the subject matter related to a given item to omit it and go on to the next by pressing a key without having

to choose any of the multiple-choice options in the earlier item. Moreover, each student was asked to cooperate with the research (Tatsuoka & Tatsuoka, 1978) of a response time model by not guessing the items. This statement appeared at each computer page in which an item was given one at a time. There were 88 students who tried every item. The fact that the test had a coefficient α_{21} reliability of .947 (Tatsuoka & Tatsuoka, 1978) lends support to our belief that guessing was indeed minimized, and hence that use of the two-parameter logistic model was justifiable.

A computer program for calibrating the parameters of this model by the MXL method was written on the PLATO system, and it was used to simultaneously obtain the values of item and ability parameters. Four items and five examinees were omitted in the calibration procedure because they caused trouble in the convergence.

With the item parameters thus obtained, various subsets were chosen from the remaining 44 items, and ability parameters of 83 examinees, based on each subset, were estimated by both the MXL and least-squares methods. Comparison of the two estimates are summarized in Table 2.

Insert Table 2 about here

The multiple Rs of the multiple regression equations of θ onto seven different sets of items are given in the second column of Table 2. The number of cases (i.e., response patterns) for which iterations for

θ did not converge are shown in columns 3 and 4 for the LS and MXL methods, respectively. The estimated θ values are obtained for all examinees by the LS method, while 12 to 43 percent of the examinees' response patterns led to nonconvergence by the MXL method. But unlike the examples for three parameter logistic model given in the previous page, the nonconvergent response patterns here are either all-right or all-wrong cases. The program that estimates ability parameter θ with known item parameters by the MXL method is a revised version of the θ estimation routine in LOGIST, modified by J.B. Sympson and adapted to the PLATO system at the University of Illinois. Since Sympson's program has been widely used in the practice of adaptive testing (Bejar & Weiss, 1978; Reckase, 1979), it is purposefully used for comparison of the two methods.

In order to measure the closeness of the $\hat{\theta}$ values estimated by the two methods, the squared differences are summed and divided by N. These values are shown in column 5 of Table 2. The correlations between the two estimated ability parameters were also calculated and are shown in column 6.

Insert Figure 1 about here

Figure 1 is a scatterplot of $\hat{\theta}$ estimated by the two methods for the subset consisting of items 12 through 21, omitting the cases where MXL did not converge. The horizontal axis represents the values of $\hat{\theta}$ obtained by the MXL method, while the vertical axis represents those estimated by the least-squares method.

Insert Table 3 about here

Since our model in this example assumed θ to be normally distributed with zero mean and unit standard deviation, the original item discriminating powers (a's) and difficulties (b's) calibrated by using the two-parameter logistic model were adjusted so that the $\hat{\theta}$ there follows the same distribution. Table 3 shows the list of adjusted a's and b's obtained from the two-parameter logistic model, and β weights, for items 12 through 21.

SUMMARY AND DISCUSSION

A latent trait variable (ability level or achievement level) $\underset{\sim}{\theta}$ was estimated by the least-squares method, in other words by regressing the variable $\underset{\sim}{\theta}$ onto a set of observed scores on given test items. It was demonstrated that the standardized β coefficient for an item entered at any stage in the multiple regression equation can be calculated from a sequence of recursively defined intervening variables α_{ik}; the earlier βs remain unchanged in the process. If θ is normally distributed, then the β_i, i=1,2, . . ., n are expressed as functions of item discriminating power $\underline{a}_i$, and the constant term is a function of the item parameters $\underline{a}$s and $\underline{b}$s (difficulties). By setting $\underline{a}_i$ to unity, the β_i, i=1,2, . . ., n for the Rasch model are obtained. More generally, since β_i is determined by replacing the correlation between $\underset{\sim}{\theta}$ and each item score by some known approximation, it is not necessary to have the item parameters obtained from ICCT models a priori. Bayesian estimates of intra-class correlation coefficients (Box & Tiao, 1975) may be substituted for $\rho_{\theta i}$ in the formula of β_i. Further research is needed along this line.

The advantage of using the least-squares method as against the MXL method may be that values of $\hat{\theta}$ are easily obtainable even for a small number of examinees and a small number of items. Mokken states in his book (1970), "In the case of Birnbaum models we first have to define by means of research the class of items which satisfies the requirements of the trace lines . . . Then the item parameters

will have to be estimated as precisely as possible before the sets of items can be used as instruments for the measurement of subjects: The estimation of θ values." The logistic models and normal ogive models that are available in practical uses at present require a restrictive assumption of unidimensionality among test items. Constructing a test satisfying the requirement for obtaining precise estimates of item parameters must be preceded by a long-term accumulation of careful efforts involving large numbers of items and subjects. Most ability tests and perhaps some achievement tests may be in this category, but many achievement tests are not, especially those used in conjunction with instruction. In these cases it may be impossible to gather the information based on a large set of items and large number of subjects in order to construct tests for which ICCT models are applicable. Yet, researchers and teachers often need to estimate students' achievement level for assessing the effect of teaching or grading test performances before enough data is gathered. The least-squares method may be a viable alternative to the maximum likelihood estimation procedure, or to Owen's restricted Bayesian updating method.

The Owen method depends on the order in which items are chosen in generating the sequence, $\{\theta_m\}$, especially when the number of items is small (Sympson, 1977). The values estimated by the least-squares method do not depend on the order of the items in a sequence but require the normality assumption of a latent

variable θ just as the Owen method does. The normality assumption of θ may not be always desirable in practical uses. Research on the robustness of the least-squares estimation procedure with respect to non-normality is necessary.

The accuracy of the estimations were compared and it was concluded that the least-squares method is just as good as the maximum likelihood method, at least for a set of the real data gathered in a matrix algebra course at the University of Illinois. However, the comparison was made between estimated θs from the least-squares and maximum likelihood methods, therefore there is no way of knowing the exact degree of accuracy of the new method in estimating the true value of a latent trait variable θ. A Monte Carlo study could provide an answer to this question.

FOOTNOTES

1. It is sometimes called a latent trait theory, or Item Response Curve Theory.

2. It is often called adaptive testing, but the author avoided its use in order to make a distinction between adaptive testing which measures ability level and diagnoses misconceptions (Brown and Burton, 1978).

Table 1

One convergent and four nonconvergent response patterns for estimating θ by the MXL method

	Response Patterns (49 items ordered roughly from easier to harder)	estimated θ	No. of iterations
1**	0000101100000000100011000000000100001000000001001	-3.798	25*
2**	1001000100001000000100000100100000010100001000000	-3.763	25
3**	1111000010000000000000000000100010000100110000000	-1.335	7
4	0000000000000000000000000000000000000011111111111	-4.957	25
5	0000000000000000000000000000111111111110000000000	-25.00	25

* Iterations were terminated at 25 tentatively, but the decrements of four cases exceeded .001.

** These three response patterns are taken from the real data and the other two are hypothetical response patterns.

TABLE 2

Comparison of Maximum Likelihood (MXL) and Least-Squares (LS) Estimation Methods in the Matrix Algebra Pretest Items (N = 83)

Items	Multiple R*	# of cases in which iterations for θ do not converge		D**	Correlation between the θs obtained by LS and MXL	Standard Deviation of s by MXL
		LS	MXL			
35-41	.970	0	24	.060	.974	.939
41-44, 47,48	.764	0	36	.102	.947	.855
12-16	.721	0	20	.153	.968	.889
31-35	.773	0	27	.064	.957	.814
12-21	.891	0	12	.069	.952	.848
18-21, 23-25, 29-31	.930	0	10	.085	.933	.817
22-31	.807	0	10	.146	.919	.904

*Multiple R is also the standard deviation of LS estimates since θ is rescaled so as to have a normal distribution with mean 0 and standard deviation one.

**The summation of the squared difference of estimated θ by the LS method from the MXL method over the number of subjects:

$$D = \sum(LS-MXL)^2/N$$

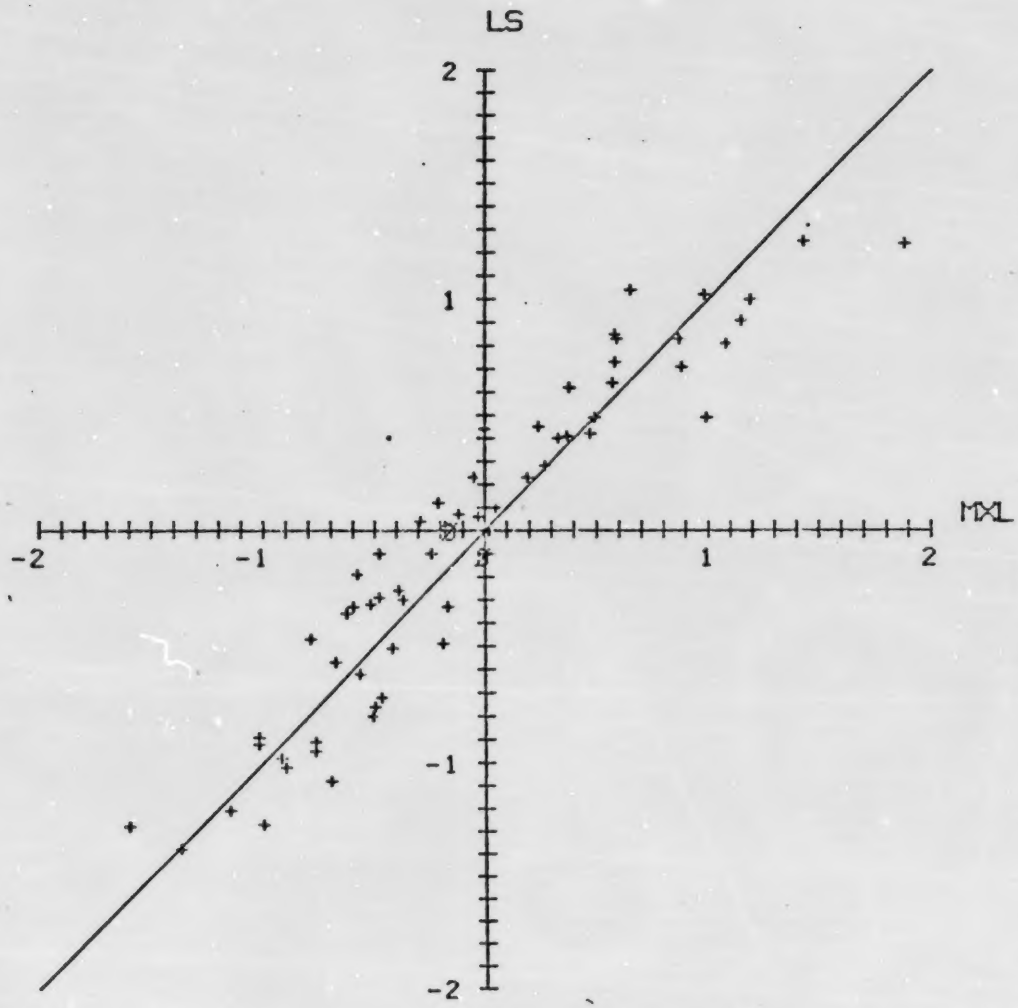

Figure 1

Comparison of θs Estimated by the MXL and LS Methods

TABLE 3

β Weights of the Least-Squares Estimation Method

Items	a^*	b^*	β
12	.269	1.049	-.014
13	.497	-.025	-.027
14	.469	-.213	.325
15	.699	-.483	.224
16	.733	-.864	.361
17	.815	.485	.409
18	.929	-.619	.736
19	.674	.485	.219
20	.712	.014	.406
21	.608	.137	.191

*a's and b's from the two-parameter logistic model were adjusted so that θ is normally distributed with zero mean and unit standard deviation.

REFERENCES

Bejar, I. I., & Weiss, D. J. A construct validation of adaptive achievement testing. (Research Report 78-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1978.

Bejar, I. I., & Weiss, D. J. Computer programs for scoring data with item characteristic curve models. (Research Report 79-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, February 1979.

Bock, R. D. Estimating item prrameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 1972, 37, 29-51.

Box, G. E. P., & Tiao, G. C. Bayesian inference in statistical analysis. Reading: Addison-Wesley, 1973.

Brown, J. S., & Burton, R. R. Diagnostic models for procedual bugs in basic mathematical skills. Cognitive Science, 1978, 2 155-192.

Brown, J. S., & Weiss, D. J. An adaptive testing strategy for achievement test batteries. (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A046062).

Dunford, N., & Schwartz, J. T. Linear Operators. (Part I: General Theory). New York: Interscience Publishers, Inc., 1964.

Jensema, C. A simple technique for estimating latent trait mental test parameters. Educational add Psychological Measurement, 1976, 36, 705-715.

Kendall, M. G., & Stuart, A. The advanced theory of statistics. New York: Itafner, 1973.

Kingsburg, G. G., & Weiss, D. J. An adaptive testing strategy for mastery decisions. (Research Report 79-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1979.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading: Addison-Wesley, 1968.

Lord, F. M. Some test theory for tailored testing. In W. H. Holtzman (Ed.), Computer-assisted instruction, testing and guidance. New York: Harper and Row, 1970.

Lord, F. M. Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters. Research Bulletin 75-33, Princeton, N. J.: Educational Testing Service, 1975.

Mokken, R. J. A theory and procedure of scale analysis. The Hague: Mouton, 1971.

Owen, R. J. A Bayesian sequential procedure for sequential response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 1979, 4(3), 207-230.

Samejima, F. Effects of individual optimization in setting the boundaries of dichotomous items on accuracy of estimation. Applied Psychological Measurement, 1978, 1, 77-94.

Simmons, G. F. Topology and modern analysis. New York: McGraw-Hill, 1963.

Sympson, J. B. Estimation of latent trait status in adaptive testing procedures. In D. J. Weiss (Ed.), Applications of computerized adaptive testing. (Research Report 77-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1977. (NTIS No. AD A038114).

Tatsuoka, K. K. Vector-geometric and Hilbert space reformation of Classical Test theory. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, 1975.

Tatsuoka, K. K., & Tatsuoka, M. M. Time-score analysis in criterion-referenced tests. (CERL Report E-1). Final Report for NIE Project No. NIE-6-76-0087. Urbana, Illinois: University of Illinois, Computer-based Education Laboratory, February 1978.

Urry, V. W. Approximations to item parameters of mental test models and their uses. Educational and Psychological Measurement, 1974, 34, 253-269.

Weiss, D. J. The stratified adaptive computerized ability test. (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Laboratory, 1973 (AD 768376).

Wood, R. L., Wingersky, M. S., & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum 76-6. Princeton, N.J.: Educational Testing Service, 1976 (revised 1978).

ACKNOWLEDGEMENT

This study was supported by the Office of Naval Research under Contract No. N00014-78-C-0159.